

CLAIMS

WHAT IS CLAIMED IS

1. A method for sorting a plurality of protein structures and corresponding sequences wherein both the protein structures and sequences are generated from a database query comprising the steps of :
 - a. identifying one or more alignment domains in each said sequence;
 - b. selecting a master sequence comprising one or more alignment domains from said sequences;
 - c. displaying to the user a graphical user interface comprising:
 - i. a first multiple sequence alignment representation comprising: a multiple sequence alignment between said master sequence and said other sequences; identifications of at least one master sequence alignment domain; and identifications of at least at least one alignment domain in each said other sequence that comprises said multiple sequence alignment;
 - ii. a means for the user to identify a new master sequence;
 - iii. a means for the user to select one or more master sequence alignment domains; and
 - iv. a means for the user to select to one or more sequences in order to select their corresponding structures for processing.
 - d. receiving at least one master sequence alignment domain selection made by a user using said means for selecting a master sequence alignment domain;

- e. displaying to the user a second multiple sequence alignment representation wherein the sequences that comprise the multiple sequence alignment are shifted such that said selected master sequence alignment domain(s) are aligned with their respective homologous alignment domains that comprise the other said sequences in said multiple sequence alignment;
 - f. receiving at least one sequence selection made by a user using said means for selecting a sequence in order to select its corresponding structure; and
 - g. identifying the protein structures corresponding to the selected sequences thereby sorting the protein structure into those selected by the user for subsequent processing and those structures that are not selected by the user for subsequent processing.
2. The method of claim 1 wherein said alignment domains correspond to domain annotations found in the Pfam database, SMART or the COG database.
3. The methods of claim 1 wherein said subsequent processing is the visualization of said protein.
4. A method for sorting a plurality of protein structures and corresponding sequences wherein both the protein structures and sequences are generated from a database query comprising the steps of :
- a. identifying one or more alignment domains in each said sequence;
 - b. selecting a master sequence comprising one or more alignment domains from said sequences;
 - c. displaying to the user,
a first graphical user interface comprising:

- i. a first multiple sequence alignment representation comprising: a multiple sequence alignment between said master sequence and said other sequences; identifications of at least one master sequence alignment domain; and identifications of at least at least one alignment domain in each said other sequence that comprises said multiple sequence alignment;
 - ii. a means for the user to identify a new master sequence;
 - iii. a means for the user to select one or more master sequence alignment domains; and
 - iv. a means for the user to select to one or more sequences in order to select their corresponding structures for processing; and
- a second graphical user interface comprising a phylogenetic tree representation of each sequence that comprises said multiple sequence alignment representation;
- d. receiving at least one master sequence alignment domain selection made by a user using said means for selecting a master sequence alignment domain;
 - e. displaying to the user via said first and second graphical user interfaces, respectively:
 - i. a second multiple sequence alignment representation wherein the sequences that comprise the multiple sequence alignment are shifted such that said selected master sequence alignment domain(s) are aligned with their respective homologous alignment

- domains that comprise the other said sequences in said multiple sequence alignment; and
- ii. for each selected master sequence alignment domain, a phylogenetic tree representation of the selected master sequence alignment domain and its homologous alignment domains that comprise the other said sequences in said multiple sequence alignment;
 - f. receiving at least one sequence selection made by a user using said means for selecting a sequence in order to select its corresponding structure for processing; and
 - g. identifying the protein structures corresponding to the selected sequences thereby sorting the protein structure into those selected by the user for subsequent processing and those structures that are not selected by the user for subsequent processing.
5. The method of claim 4 wherein said alignment domains correspond to domain annotations found in the Pfam database, SMART database or the COG database.
6. The methods of claim 4 wherein said subsequent processing is the visualization of said protein.
7. A method for sorting a plurality of protein structures and corresponding sequences wherein both the protein structures and sequences are generated from a database query comprising the steps of :
- a. identifying one or more alignment domain in each said sequence;

- b. selecting a master sequence comprising one or more alignment domains from said sequences;
- c. displaying to the user,
 - a first graphical user interface comprising:
 - i. a first multiple sequence alignment representation comprising: a multiple sequence alignment between said master sequence and said other sequences; identifications of at least one master sequence alignment domain; and identifications of at least at least one alignment domain in each said other sequence that comprises said multiple sequence alignment;
 - ii. a means for the user to identify a new master sequence;
 - iii. a means for the user to select one or more master sequence alignment domains; and
 - iv. a means for the user to select to one or more sequences in order to select their corresponding structures for processing;
 - a second graphical user interface comprising a phylogenetic tree representation of each sequence that comprises the multiple sequence alignment representation; and
 - a third graphical user interface comprising a table consisting of data fields for the source and name of each corresponding sequence that comprises the multiple sequence alignment;

- d. receiving at least one master sequence alignment domain selection made by a user using said means for selecting a master sequence alignment domain;
- e. displaying to the user via said first, second and third graphical user interfaces, respectively:
 - i. a second multiple sequence alignment representation wherein the sequences that comprise the multiple sequence alignment are shifted such that said selected master sequence alignment domain(s) are aligned with their respective homologous alignment domains that comprise the other said sequences in said multiple sequence alignment;
 - ii. for each selected master sequence alignment domain, a phylogenetic tree representation of the selected master sequence alignment domain and its homologous alignment domains that comprise the other said sequences in said multiple sequence alignment; and
 - iii. a table consisting of data fields for the source and name of each corresponding sequence that comprises the multiple sequence alignment;
- f. receiving at least one sequence selection made by a user using said means for selecting a sequence in order to select its corresponding structure for processing; and

- g. identifying the protein structures corresponding to the selected sequences thereby sorting the protein structure into those selected by the user for subsequent processing and those structures that are not selected by the user for subsequent processing.
- 8. The method of claim 7 wherein said alignment domains correspond to domain annotations found in the Pfam database, SMART database or the COG database.
- 9. The methods of claim 7 wherein said subsequent processing is the visualization of said protein.
- 10. The method of claim 10 wherein said table further comprises data fields for:
 - a. a database annotation score that reflects the relative amount of information known about said sequence;
 - b. a sequence similarity score that reflects the evolutionary similarity of said sequence with said master sequence; and
 - c. a sequence identification number that identifies said sequence from a plurality of sequences in a database.
- 11. A computer system comprising:
 - a. a processor;
 - b. an input means;
 - c. an output means;
 - d. a memory
 - e. an operating system means;

f. programming for displaying to the user a graphical user interface

comprising:

- i. a multiple sequence alignment representation comprising: a multiple sequence alignment between a master sequence and a plurality of other sequences; identifications of at least one master sequence alignment domain; and identifications of at least at least one alignment domain in each said other sequence that comprises said multiple sequence alignment;
- ii. a means for the user to identify a new master sequence;
- iii. a means for the user to select one or more master sequence alignment domains; and
- iv. a means for the user to select to one or more sequences in order to select their corresponding structures for processing.

12. The system of claim 11 wherein said means for the user to select one or more master sequence alignment domains comprises displaying to the user a selectable check box corresponding to each master sequence alignment domain.

13. A computer system comprising:

- a. a processor;
- b. an input means;
- c. an output means;
- d. a memory
- e. an operating system means;
- f. programming for displaying to the user simultaneously,

a first graphical user interface comprising

- i. a multiple sequence alignment representation comprising: a multiple sequence alignment between a master sequence and a plurality of other sequences; identifications of at least one master sequence alignment domain; and identifications of at least at least one alignment domain in each said other sequence that comprises said multiple sequence alignment;
- ii. a means for the user to identify a new master sequence;
- iii. a means for the user to select one or more master sequence alignment domains; and
- iv. a means for the user to select to one or more sequences in order to select their corresponding structures for processing; and

a second graphical user interface comprising, for each selected master sequenced alignment domain, a phylogenetic tree representation between a master sequence alignment domain and a plurality of alignment domains homologous to said master sequence alignment domain.

14. The system of claim 13 wherein said means for the user to select one or more master sequence alignment domains comprises displaying to the user a selectable check box corresponding to each master sequence alignment domain.

15. A computer system comprising:

- a. a processor;
- b. an input means;
- c. a memory;

- d. an output means;
- e. an operating system means;
- f. programming for displaying to the user simultaneously,
 - a first graphical user interface comprising:
 - i. a multiple sequence alignment representation comprising: a multiple sequence alignment between a master sequence and a plurality of other sequences; identifications of at least one master sequence alignment domain; and identifications of at least at least one alignment domain in each said other sequence that comprises said multiple sequence alignment;
 - ii. a means for the user to identify a new master sequence;
 - iii. a means for the user to select one or more master sequence alignment domains; and
 - iv. a means for the user to select to one or more sequences in order to select their corresponding structures for processing;
 - a second graphical user interface comprising, for each selected master sequenced alignment domain, a phylogenetic tree representation between a master sequence alignment domain and a plurality of alignment domains homologous to said master sequence alignment domain; and
 - a third graphical user interface comprising a table consisting of data fields for the source and name of each corresponding sequence that comprises the multiple sequence alignment.

16. The system of claim 15 wherein said means for the user to select one or more master sequence alignment domains comprises displaying to the user a selectable check box corresponding to each master sequence alignment domain.

17. A computer system comprising:

- a. a processor;
- b. a memory;
- c. an input means;
- d. an output means;
- e. an operating system means;
- f. programming for the methods according to claim 3; and
- g. programming for displaying protein structures based upon their structural coordinates

18. A computer system comprising:

- b. a processor;
- c. a memory;
- d. an input means;
- e. an output means;
- f. an operating system means;
- g. programming for the methods according to claim 6; and
- h. programming for displaying protein structures based upon their structural coordinates

19. A computer system comprising:

- a. a processor;

- b. a memory;
- c. an input means;
- d. an output means;
- e. an operating system means;
- f. programming for the methods according to claim 9; and
- g. programming for displaying protein structures based upon their structural coordinates.